Gu Test: A Progressive Measurement Of Generic Artificial Intelligence

Author: Lifan Gu□□□□□

Turing Test is subjective. It is an empirical test, not a scientific experiment. Langauge complexity is much less than the human intelligence complexity. So Turing Test is invalid.

Sciences are different from mathematics. Scientific experiements only can falsify, but never prove unlimited possiblities. Scientific research is an ongoing process, should always open to new experiments.

So existing empirical tests for artificial intelligence (AI) technologies, such as the regular Go games played by AlphaGo Zero and other computer Go systems, the simulations and road tests for self-driving cars, the datasets for natural langauge understanding, etc. are also inadequate.

Technological Singularity is baseless. Driverless cars with no constrains (i.e. SAE level 5 automated-driving) are impossible. There are problems in the definition of SAE level 4.

In reality, there is no way to prove a car with SAE level 4 automated-driving ability, especially when the mode evolution in future is not stable. So new concepts of AI and new definitions of automated-driving should be studied.

In this paper, I will discuss the problems in Turing Test, the problems in existing testing of AlphoGo Zero, self-driving cars, natural langauge understanding, and the problems in the mainstream textbook AI: A Modern Approach. Then I will propose Gu Test, a progressive measurement of generic artificial intelligence, based on falsifiability, which could help to develop scientific intelligence theories gradually.

1. The Problems in Turing Test

Turing Test is invalid, but still cause misleading widely in AI research so far.

Many existing tests for AI techonologies have similar problems as Turing Test. So it is important to analyze its problems and clarify the misleading.

Turing Test is subjective. Testing it with different people could yield very different results. People with different knowledges, especially with different understanding levels of computer technologies, could give very different results. The subjectiveness of Turing Test cause unstable results, which makes Turing Test invalid.

Moreove, language complexity is much less than human intelligence complexity. Humans have much more intelligence than language level intelligence [1]. So Turing Test is not valid by making judgement of intelligence based on language conversation. Indistinguishablity between humans and computers by language conversations does not mean equivalence of intelligence.

Turing Test is also an empirical test, not a scientific experiment.

Sciences are different from mathematics. Scientific experiements only can falsify, but can never prove unlimited possiblities. Actually, equivalence of intelligence between humans and computers can never be proved, but only can be falsified.

Scientific research is an ongoing process, should always open to new experiments. If computers pass some tests, other people still could design new tests to disprove.

Scientific experiments should be done with strictly controlled conditions, to test the underlying principles. Scientific conclusions can only be derived from these principles based on the strict conditions. From empirical tests, people can not derive scientific conclusion.

Other existing tests for AI technologies have many similar problems. In the next sections, I will discuss the testing problems for computer Go systems, self-driving cars, and natural langauge understanding.

2. AlphaGo Zero's Superhuman Claim

The AlphaGo Zero paper on Nature magazine [2] claimed superhuman performance. However it did not provide any evidences for this claim, did not provide any evidences to show AlphaGo Zero is superior to GENERIC

human even in Go gaming.

AlphaGo Zero defeated AlphaGo Master and other AlphaGo series programs is not an evidence of superhuman, because these computer Go systems also suffer from the limitations of AI.

Even if AlphaGo Zero could defeat all human players in regular Go games, this still does not provide superhuman evidences, because these human players still do not know the limitations of AlphaGo Zero and AI yet.

As long as someone could design some scientific experiments with strictly controlled conditions, to let human players know the weakness of AlphaGo Zero and AI [3], human players still could defeat AlphaGo Zero and other computer Go systems in fair Go games.

Scientific experiments are different from regular gaming. Regular gaming is just to win a game or some games. Scientific experiments is to falsify some assumptions or let the assumptions pass the experiments. Scientific research should always open to new experiments.

Scientific experiments should be done with strictly controlled conditions. The conclusions of the experiments only can be derived from the controlled conditions and the results.

Scientific experiment results should open to discussions, so other people could discuss whether the experiments are valid and whether the interpretations of these results are justified or not.

The Go games played by AlphaGo Zero and other computer Go systems are just regular GAMES or empirical tests, not scientific experiments. People cannot derive scientific conclusions from these games or empirical tests.

So the superhuman claim from the AlphaGo Zero paper on Nature magazine is not a scientific conclusion. Actually, as already analyzed, this superhuman claim is baseless.

Go is a game with simple rules and good abstraction, but still significant complexity, which make it an ideal tool to design scientific experiments [4].

It is much easier to isolate various factors and figure out the intelligence principles in such simpler experiments with strictly controlled conditions.

3. Test Automated Driving

The current simulations and road tests for self-driving cars are not scientific experiments. They are just empirical tests without strictly controlled conditions. So from these simulations and road tests, we cannot derived any scientific conclusions [5].

Driverless cars with no constrains (i.e. SAE level 5 automated-driving) are impossible, which could be verified by scientifc experiments with the AI technologies used by these self-driving cars.

There are problems in the definition of SAE level 4. In different areas, there are very different requirements for SAE level 4 automated-driving techniques. Even in the same areas, if the mode evolution in future is not stable, severe problems could appear with large probabilities even if the cars already passed SAE level 4 in previous modes.

Scientifc experiments with the AI technologies used by self-driving cars could verify that when the mode evolution is not stable, severe problems with large probabilities could occur in the systems with very little problems in previous modes. So in reality, there is no way to prove a car has truly passed SAE level 4 automated-driving.

Although people could solve some problems in self-driving cars case by case, if they do not know the underlying principles, they would not solve certain important root causes of the problems. These root causes could appear in very different forms in future, especially if the mode evolution is not stable [6].

To understand unstable mode evolution, we need understand the underlying principles of intelligence in these AI technologies first. So it is a luck that we could study the AI problems of with theoretic analyses, and design experiments based on these analyses, to verify the AI problems with simpler systems, such as AlphaGo Zero.

I wrote before: Go gaming is strictly defined within a very small space. Industrial automations are typically designed in environments well controlled, but not strictly defined. Car driving is regulated, but the environment is not well controlled.

Many industrial automation problems can not be solved yet. As said, there are still serious problems even in AlphaGo Zero which could be verified by scientific experiments.

Based on my previous analyses and my experiment plans, I have reasons to believe: many years after large-scale deployment of self-driving cars, regular people would have enough chance to interact with self-driving cars from different aspects and trigger unstable mode evolutions which cause severe problems with large probabilities. However, at that time, it would be too late.

The technologies for traditional automobiles, such as electronics, powertrain, and other mechanics, etc. are based on concrete sciences whose main principles are already well tested in sciences. However, AI for automated driving is empirical, with no scientific foundation. It could be very unstable in future mode evolution. So testing automated driving vehicles in a similar way to testing traditional vehicles is very misleading.

I do not study automated driving directly myself. It is better to do foundamental studies first and figure out the underlying principles of intelligence. By verifying these principles and the problems in AI with simpler and undangerous AI systems, we could avoid the problems of deploying unmature self-driving cars in large scale.

4. The Problems in AI: A Modern Approach

5. Measure Language Intelligence

AI could do searches well and have a much better memory for text contents than humans. AI even could achieve many progresses in machine translation. However, AI does not really understand semantics. There is a Chinese room issue, which could be verified.

AI could not process high-order logic properly, could not recognize sophism, could not recognize wrong thinking modes, such as Aristotle thinking mode.

So replying on AI to make judgement could cause severe problems in juridical practice, scientific researches, education, medical practice, etc. Asking students to obey computer's thinking mode could damage their intelligence development.

The current testing datasets for language understanding, such as SQuAD, CoQA, QuAC, NLVR^2, GLUE series, cannot measure the real difference between human and Natural Language Processing (NLP). They cannot help much on high-order logic processing, recognizing sophism, verify Chinese room issues, etc.

All of these datasets fall into the traps of Aristotle thinking mode. They can not recognize wrong thinking modes, and are not scientific methods.

To understand human intelligence, we need a structural and systematic analysis of human intelligence. I defined certain main intelligence levels: language level, philosophical level, mathematical level, scientific level, all with different requirements and criteria.

Langauge intelligence is an important characteristic of human intelligence. Other known lives do not have advanced language ability. Langauge is also an important media for human knowledge, the basis for philosophy, mathematics, sciences, etc.

Based on languages, humans developed two important branches of studies: mathematics and philosophy. Mathematics develops towards accuracy. Philosophy develops towards integrity.

Sciences originates from philosophy, so sciences also develop towards integrity. More than philosophy, sciences make conclusions based on experiments of falsifiability with strictly controlled conditions. Beyond philosophy, sciences also gradually introduce accuracy and mathematics.

Mathematics does not meet the criteria of sciences. It even does not have integrity [7].

Based on these structural and systematic studies of human intelligence, people could measure langauge intelligence much better.

6. Gu Test

7. Conclusion

---

[1] In section 5., I will discuss more on different intelligence levels.

[2] Mastering the game of Go without human knowledge, Published: 18 October 2017:

https://www.nature.com/articles/nature24270

[3] Actually I designed such experiments as introduced in the section 6 of this article, and requested Deepmind to do the experiments, but they have

not accepted the experiments.

Scientific research should be based on open discussion and fair experiment. So the superhuman claim for AlphaGo Zero is not a scientific conclusion.

[4] I began to consider to use computer Go systems to measure AI technologies long before Deepmind started AlphaGo project.

[5] According to some news, in 2015 a blind man was allowed to take a driverless car alone, before the accident on 02/14/2016. Although the damage of this accident is minor, wrong judgment of driverless cars is very dangerous potentially, especially if the mode evolution in future is unstable.

"Steve Mahan, who is legally blind, was the first non-Google employee to ride alone in the company's gumdrop-shaped autonomous car. The ride was in October 2015 in Austin. (Courtesy Waymo)", https://www.washingtonpost.com/local/trafficandcommuting/blind-man-sets-out-alone-in-googles-driverless-car/2016/12/13/f523ef42-c13d-11e6-8422-eac61c0ef74d_story.html,

"Steve Mahan, who is legally blind, takes what Waymo called the world's first fully autonomous ride in Austin in 2015, in an image provided by the Alphabet unit.", https://www.marketwatch.com/story/google-says-driverless-cars-are-ready-to-make-money-but-we-wont-know-if-they-do-2016-12-13.

[6] In a MIT lecture published on Feb 12, 2019, Drago Anguelov. a Principal Scientist at Waymo, admitted that there is a long tail of problems in self-driving cars: https://www.youtube.com/watch?v=Q0nGo2-y0xY.

The real situation could be more complicated than a long tail. The underlying principles of intelligence could help us to understand how the problems evolve, and transform, etc.

[7] For more details, please see my article: A Structural and Systematic Analysis of Human Knowledge and Studies.